IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF NEW YORK

UNITED STATES OF AMERICA	)	
	)	
v.	)	No. 23 Cr. 13 (NRM)
	)	
TYLER SCOTT JOHNSTON,	)	
Defendant.	)	

DEFENDANT'S MOTION UNDER FEDERAL RULES OF EVIDENCE 401, 403, 702, AND *DAUBERT* TO PRECLUDE DNA TESTIMONY

Samuel I. Jacobson
Marissa Sherman
Deirdre D. von Dornum
Austin Dean (Legal Intern)
Federal Defenders of New York
Counsel for Tyler Johnston

TABLE OF CONTENTS

EXHIBIT LIST

TABLE OF AUTHORITIES

Cases

Rules

Other Authorities

Defendant Tyler Johnston respectfully moves under Federal Rules of Evidence 401, 403, 702 and *Daubert v. Merrell Dow Pharms., Inc*., 509 U.S. 579, 589 (1993) to preclude the government from offering at trial the opinion evidence proffered by a U.S. Army Forensic Biologist analyzing material found on a comforter in the defendant and his wife's bedroom.

MEMORANDUM OF LAW

I. PRELIMINARY STATEMENT

This case presents a worst-case scenario for forensic biological analysis: several complex mixtures of DNA were collected from a marital bed that reasonably could be expected to include background contributions from six first-order relatives, including the defendant, his wife, and the complainant child. The case allegations, which were known to the Biologist as she conducted her analyses, involve sexual abuse of a child where the child's accusations are uncorroborated by any other evidence. Further, the serological testing conducted to search for evidence of sexual contact failed to identify sperm, and every sample analyzed was determined to include at least three contributors, if not more. Under these circumstances—where all persons of interest are *expected* contributors, kinship is complicating the data, and there is no concrete evidence corroborating sexual contact—the forensic biological data becomes both meaningless and extremely dangerous.

The government seeks to put before the jury the testimony of a U.S. Army analyst, Sara Green, that there were several semen stains on a comforter recovered from the bed, and that the father, the complainant daughter, and her mother were "identified as contributors" to these stains. *See* Expert Disclosure of Sara Green, attached as Exhibit A. This opinion testimony, however, stems not from the data, but from the analyst's threshold assumptions. Three misguided, compounding assumptions underlie the proffered opinion testimony:

(1) Several of the stains on the comforter were identified as semen, as if they could not be any other biological or nonbiological material;
(2) Almost all of the stains on the comforter were assumed to contain a mixture of just three contributors; and
(3) Because the stains were assumed to be semen and to contain the DNA of three contributors, it was assumed that only the defendant, his wife, and the complainant child could be contributors to those stains. The other three children in the home—whose DNA profiles are unknown—were inferentially excluded.

To reach the categorical conclusion that some of the stains on the comforter were in fact semen, Ms. Green had to assume that they were. No sperm were visualized in the slides Ms. Green prepared, and the samples that proceeded to differentially extracted DNA testing contained no DNA in the sperm fraction. Both serological testing methods Ms. Green used to attempt to identify semen are accompanied by significant testing uncertainty. As the recent Report of the Expert Working Group on Human Factors in Forensic DNA Interpretation warned regarding serological test result reporting: "The use of phrases regarding a test's ability to identify or confirm a biological material implies a level of certainty that the testing methods cannot support while also ignoring the potential for false positives and false negatives." *Forensic DNA*

Interpretation and Human Factors: Improving Practice Through a Systems Approach (2024), at 108 (hereinafter "*Human Factors*"), attached as Exhibit C. Ms. Green's analysis, thus, unacceptably failed to account for the possibility that at least some of the initial serological screening tests were positive not because of semen but because of other biological or nonbiological materials. Further, her proffered opinion conveys a level of certainty that is not scientifically supported and has the strong potential to mislead the jury.

To reach the second categorical conclusion, that those stains were from just three contributors, Ms. Green had to assume that they were. Because her laboratory's internal validation had not considered the problem of kinship and high allele sharing, Ms. Green failed to consider the complexity introduced by the presence of multiple first-order relatives in the environment. Imagine: there were four people in that household that share at least half their DNA with the defendant's wife, two that share at least half with the defendant, two that share *all* of their DNA with the combination of the defendant with his wife, and four that share somewhere between 25% and 60% of their alleles with each other.[1] As the National Institute of Standards and Technology has noted, "[w]hen alleles overlap and are shared between contributors, it becomes more difficult to definitively estimate the number of donors to the DNA mixture." Butler et al., *DNA Mixture Interpretation: A NIST Scientific Foundation Review* (2021 Draft), attached as Exhibit BB, at 46. Despite high allele sharing being among the known fundamental challenges in DNA mixture interpretation, Ms. Green never checked her assumptions about the number of contributors to each stain. *See id.* at 30 ("There are at least three challenges that are fundamental to DNA mixture interpretation… (3) sharing of common alleles, which influences the ability to estimate the number of contributors…."). The lack of internal validation studies examining the problem of kinship and the determination of the number of contributors in the face of kinship, the lack of clear guidance in the Army Lab's protocols, and Ms. Green's own failure to recognize the intricacy presented by the circumstances resulted in an unreliable opinion as to the number of contributors in the stains.

To reach the third, related, categorical conclusion, that those stains were a combination of the defendant, his wife, and his complainant daughter, Ms. Green had to maintain her assumptions about semen and number of contributors. She then analyzed each stain in light of the profiles of the defendant, his wife, and his complainant daughter, and failed to account for the potential that any of the other children in the home might be present in the samples, that any of the children in the home might trigger a false positive likelihood ratio for the defendant or for his wife, or that any other unknown contributors to the mixtures—related or unrelated—might be possible explanations for the unconditioned data. In other words, Ms. Green painted the bullseye around her target.

[1] This comparison, while staggering, fails to account for the exponential explosion in difficulty that arises from the cross-combination of these genomes as the number of contributors to any mixed sample increases. An example: it is possible that all the defendant's alleles would be masked in a mixture of his wife and his two biological children. Similarly, it is probable that the defendant's wife's alleles would all be masked in a mixture of the four children in the home. And so on.

Under these circumstances, the opinion testimony will not assist the trier of fact, is not based on sufficient facts or data, and is not the product of reliable principles and methods.

The DNA analysis the government seeks to admit should be excluded for at least four specific reasons. First, and perhaps most fundamentally, because the tested item was used in the regular course of household activity by both the complainant and the defendant, as well as the rest of the family, the analyst's conclusion, even were it reliable, is far more prejudicial than probative, and thus cannot satisfy Federal Rule of Evidence 403. Second, at the time of the analysis, the Army Laboratory had not validated its DNA protocols to take account of kinship-based propositions. Therefore, the government cannot show that the lab's analysis of DNA mixtures would be reliable in any situation where there are multiple potential contributors, all of whom are related, but some of whose profiles are unknown. Third, and relatedly, the Army Biologist could not reliably determine the number of contributors to the DNA samples because allele sharing among family members complicates the inherently subjective number-of-contributors assessment. Because each conclusion here relied on an assumption that there were only three contributors, when the facts make clear there easily could have been four, five or six related contributors, the downstream interpretation conducted by STRmix is not reliable. Fourth, because the government cannot reliably establish that these mixtures contain fewer than five contributors, the STRmix testing protocol used by the lab in this case, which had not been internally validated for mixtures of five or more contributors, cannot be shown to be reliable.

II. LEGAL STANDARD

Federal Rule of Evidence 702 allows admission of an expert witness's testimony only if the proponent demonstrates to the court that it is more likely than not that: (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert's opinion reflects a reliable application of the principles and methods to the facts of the case. Fed. R. Evid. 702. Importantly, "[w]hile the proponent of expert testimony has the burden of establishing by a preponderance of the evidence that the admissibility requirements of Rule 702 are satisfied, the district court is the ultimate 'gatekeeper,'" and must determine that the expert's testimony "both rests on a reliable foundation and is relevant to the task at hand." *United States v. Williams,* 506 F.3d 151, 160 (2d Cir. 2007) (citations omitted).

As a threshold matter, the proponent of the expert testimony must provide a thorough disclosure of the anticipated testimony in order to enable the opposing party and the Court to evaluate the reliability of the testimony. Specifically, Federal Rule of Criminal Procedure 16, as amended in 2022, requires the proponent of an expert witness, here the government, to disclose "a complete statement of all opinions that the government will elicit from the witness" and "the bases and reasons" for these opinions. The amendments to Rule 16 were explicitly designed to address "the lack of adequate specificity regarding what information must be disclosed." *United States v. Mrabet*, No. 23 Cr. 69 (JSR), 2023 WL 8179685, at *2 (S.D.N.Y. Nov. 27, 2023) (quoting Rule 16, Notes of Advisory Committee on 2022 Amendment). Based on the amended rule, it is not sufficient for the government to provide the defendant with only "broad statement[s]" and "broadly and briefly described categories" of proffered testimony. *Id.* Failure

to provide the required information can result in the Court limiting or excluding the proffered testimony.

Once an adequate disclosure has been made, the Court undertakes its gatekeeper function. In *Daubert*, the Supreme Court "enumerated a list of factors that, while not constituting a 'definitive checklist or a test,' a district court might consider in evaluating whether a proffered expert opinion has the required indicia of scientific reliability" under Rule 702. *Nimely v. City of New York,* 414 F.3d 381, 396 (2d Cir. 2005) (quoting *Daubert,* 509 U.S. at 593–94). Those factors include whether the theory or technique (1) can be and has been tested; (2) has been reviewed by peers and published; (3) has a known or potential error rate; (4) has standards controlling the technique's operation that are maintained; and (5) is generally accepted as reliable within the relevant scientific community. *Kumho Tire Co. v. Carmichael,* 526 U.S. 137, 149–50 (1999). Beyond the core *Daubert* factors, "when an expert opinion is based on data, a methodology, or studies that are simply inadequate to support the conclusions reached, *Daubert* and Rule 702 mandate the exclusion of that unreliable opinion testimony." *Nimely*, 414 F.3d at 396–97 (quotation marks and citation omitted).

One requirement that ensures a sufficient nexus between an expert's opinion and the data, methodology, or studies supporting that conclusion is the completion of appropriate validation studies. Indeed, before conclusions grounded in a particular methodology can be admitted into evidence, the methodology must be supported by "appropriate validation." *Daubert*, 509 U.S. at 590 ("Proposed testimony must be supported by appropriate validation – i.e., 'good grounds,' based on what is known."). This principle applies with full force to probabilistic genotyping software programs like STRmix, the program used by the Army lab here. The President's Council of Advisors on Science and Technology ("PCAST") specifically advises that "[w]hen considering the admissibility of testimony about complex mixtures (or complex samples), judges should ascertain whether the published validation studies adequately address the nature of the sample being analyzed (e.g., DNA quantity and quality, number of contributors, and mixture proportion for the person of interest)." PCAST, *An Addendum to the PCAST Report on Forensic Science in Criminal Courts* 9 (2017); *accord* K. Kwong, *The Algorithm Says You Did It: The Use of Black Box Algorithms to Analyze Complex DNA Evidence*, 31 HARV. J. L. & TECH. 275, 277–82, 300 (2017) ("[C]ourts should rigorously examine whether a given algorithmic system has been validated for a particular type of evidence analysis and refuse to admit evidence that lacks demonstrated validity for a given mixture type[.]"), attached as Exhibit E.

"In addition to the requirements of Rule 702, expert testimony is subject to Rule 403." *Nimely,* 414 F.3d at 397. Accordingly, such evidence may be excluded "if its probative value is substantially outweighed by a danger of . . . unfair prejudice, confusing the issues, [or] misleading the jury…." Fed. R. Evid. 403.

A court's obligation to ensure that the proponent of expert testimony has met its burden does not evaporate when that testimony concerns the well-regarded science of DNA analysis. As the New York Court of Appeals recently emphasized,

> In the criminal justice system, [genetic biology] has provided forensic science with one of the most powerful tools for identification yet seen. DNA testing has

> become the 'gold standard' of this process. For this reason, more than any other, courts must use the tools available to make sure the highest standards of reliability are maintained.

People v. Williams, 35 N.Y.3d 24, 29 (2020).

This is particularly so because, despite a common assumption that DNA evidence is both objective and infallible, inherent subjectivity pervades "even the most favorable conditions of forensic DNA typing." E. Murphy, *The Art in the Science of DNA: A Layperson's Guide to the Subjectivity Inherent in Forensic DNA Typing*, 58 EMORY L. J. 489, 509 (2008). The complexity of forensic biological evidence, its ability to overwhelm jurors' common sense, and the high stakes risk of jurors failing to detect error makes the trial court's gatekeeping role an essential one. *See*, *e.g.*, Executive Office of the President, PCAST, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* 45 (2016) ("The vast majority of jurors have no independent ability to interpret the probative value of results," and "[t]he potential prejudicial impact is unusually high, because jurors are likely to overestimate [that] probative value…."), available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

III. DNA TESTING AND INTERPRETATION

Deoxyribonucleic acid, or DNA, is a double-stranded molecule that coils to form the characteristic double helix, and is found in all cells possessing a nucleus.[2] JOHN BUTLER, FUNDAMENTALS OF FORENSIC DNA TYPING 19 (2010) ("*Fundamentals*").[3] Forensic DNA typing examines certain locations, or loci (singular: locus), on the DNA strand. The DNA typing technique at issue in this case is short tandem repeat (STR) testing. STR typing measures how many times a short piece of DNA repeats at each of the tested loci; the number of repeats is known as an allele. *Id* at 148. An individual's genetic type, or profile, is the compilation of his or her alleles at each locus tested. At each locus, an individual possesses two alleles: one allele inherited from each biological parent. *Id* at 25. Thus, an individual's DNA profile is simply a list of two numbers—known as a ***genotype***—per locus examined. An individual can inherit the same allele—*i.e.*, the same number of repeats—at a locus from both their biological parents (*e.g.*, 12, 12). Alternatively, an individual can inherit two different alleles—two different numbers—at a locus from their parents (*e.g.*, 12, 16).

The DNA testing process proceeds via a series of standard steps: extraction, quantitation, amplification, genetic analysis, and interpretation. The first step in generating a DNA profile from a sample is extraction, where the analyst attempts to isolate the DNA and separate it from

---

[2] Most cells, with the exception of red blood cells, possess nuclei. When it is in the nucleus, DNA is tightly packaged into two sets of 23 chromosomes; one set of 23 chromosomes is inherited from each parent. Sperm and egg cells possess only one set of 23 chromosomes each; when they unite, the resulting embryo possesses the full set of 46 chromosomes. *Fundamentals* at 23.

[3] John Butler's textbooks on DNA analysis—including FUNDAMENTALS OF FORENSIC DNA TYPING and ADVANCED TOPICS IN FORENSIC DNA TYPING: INTERPRETATION—are considered authoritative in the field of forensic DNA analysis. They are used in forensic science education and are cited in laboratories' protocols.

all other cellular material and debris. *Id.* at 99. After extraction of the DNA, the sample is quantitated, *i.e.*, the total amount of DNA present in the sample is estimated. *Id*. at 114. Based on the estimated amount of DNA present, some portion of the extracted DNA is then amplified. Amplification is a process by which DNA is copied at targeted locations (called loci) many times over, generating on the order of a billion copies.[4] *Id*. at 125-26. During the amplification process, the targeted DNA may not amplify for a number of reasons, including a) if there is only a small amount of it present to begin with,[5] or b) if it is degraded (*e.g*., broken into pieces due to environmental exposure or other stressors), or c) if there are inhibitors (such as some fabric dyes or excess salts) present in the sample. *Id*. at 68. When targeted DNA does not amplify, that genetic information is lost in downstream steps; this loss of genetic information is known as allelic dropout. *Id*. at 222. The term "allelic dropout" specifically refers to a scenario in which only one of a DNA contributor's two alleles at a given locus is detected by the DNA typing process.[6]

The post-amplification sample consists of large numbers of only the copied alleles, which can then be separated on an instrument called a genetic analyzer so that each allele can be distinguished and then recorded. *Id*. at 175. The result of this process is a series of peaks on a graph, called an electropherogram. *Id*. at 194. Under the U.S. Army Lab's protocols, the next step, interpretation, is undertaken partially by an analyst and partially by proprietary software. *See* Exhibit F, DFSC DNA 114.1: Interpretation of DNA Results Using Globalfiler. For the analyst's part, she remains responsible for determining whether she is dealing with a single-source sample or a mixture and then whether peaks on the produced graph represent "real" DNA or artifacts of the testing process. Each "real" DNA peak corresponds to an allele present in the sample and the height of each peak roughly corresponds to how much of that allele is present (*i.e.*, a taller peak indicates more of a particular allele present). When testing a single source evidence sample (*i.e.*, a sample originating from one individual), two peaks of roughly equivalent height should be observed at each locus where the contributing individual possesses two different alleles. At loci where the contributor possesses two of the same allele, one relatively high peak should be observed, because the two alleles "stack" on top of one another. *See*, *e.g.*, Figure 1, below.

[4] Amplification is conducted via a technique called polymerase chain reaction, commonly abbreviated as PCR.

[5] Quantitation gives a preliminary estimate of whether the amount of DNA in the extract falls into this low-level range. However, a seemingly sufficient total amount of DNA may be comprised of low levels of DNA from multiple contributors; this is not something that can be discerned from the quantitation step, which does not distinguish between contributors but rather reports the total amount of DNA present.

[6] There can also be loss of both alleles at a given location, which is called "locus dropout."

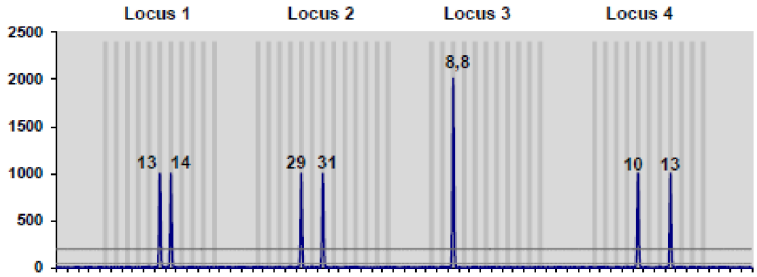

Figure 1. Electropherogram showing ideal, single-source DNA data at four hypothetical loci. Note that at Locus 3, where the DNA contributor is a homozygote (possesses two "8" alleles), his two alleles "stack" on top of one another, resulting in a single peak on the electropherogram. At each of the other three loci, where the contributor is a heterozygote (possesses two different alleles), two peaks are observed. Figure from Butler, *Advanced Topics in Forensic DNA Typing: Interpretation*, 11, Fig. 1.5 (2014).

Forensic DNA samples often contain DNA from more than one individual. Such a DNA profile representing two or more contributors is termed a DNA mixture. *Fundamentals* at 320. An analyst knows that a sample is a DNA mixture, rather than from a single source, if the analyst observes more than two alleles at two or more loci, or if loci with only two alleles display significant peak height imbalance.[7] *Interpretation*, *at* 129. Unlike the straightforward analysis involved in interpreting a high-quality, single-source DNA profile, mixtures are often ambiguous, and the process of interpreting them can be highly subjective.

Studies have shown that attempts at interpretation of such complex DNA data can lead to widely divergent results from one analyst to the next, even among analysts in the same laboratory applying the same set of protocols. *See* I. Dror & G. Hampikian, *Subjectivity and bias in forensic DNA mixture interpretation*, 51 SCI. & JUST. 204 (2011);[8] M. Coble, *MIX13: An interlaboratory study on the present state of DNA mixture interpretation in the U.S.* (2014) (available at https://strbase-archive.nist.gov/pub_pres/Coble-ABA2014-MIX13.pdf);[9] and

---

[7] Two alleles from the same contributor should be roughly the same height, within a degree of tolerance (called a "peak height ratio" (PHR)). If the height of two allelic peaks observed at a given locus are not within this predetermined tolerance—*i.e.*, they are "imbalanced"—this indicates that the peaks actually originate from two people rather than one.

[8] In this study, 17 examiners from one government laboratory were provided a mixed DNA profile from a sex assault case and asked to interpret the profile and compare it to a suspect's reference profile. The original case analyst had determined that the suspect could not be excluded as a contributor to the mixture. The 17 examiners came to a variety of conclusions: 1 concluded "cannot exclude," 12 "excluded," and 4 deemed the results "inconclusive." These results underscore the subjectivity of complex mixture interpretation.

[9] The NIST MIX13 study was the largest study of its kind, broadly assessing the accuracy, reproducibility, and repeatability of mixture interpretations among and across laboratories. Analysts from 108 laboratories took part, and 46 states had at least one laboratory participate; the participants were from a mix of federal, state, and local labs. As one of the study's leading

forerunner NIST mixture studies (*e.g.*, MIX05).

One of the biggest complicating factors in the interpretation of complex mixture data is the potential for allele sharing. *Interpretation* at 153. As described above, when an individual has two of the same allele at a locus (*i.e.*, is a homozygote), that person's alleles "stack" on top of one another and present as a single peak on the electropherogram. Similarly, when multiple contributors to a DNA mixture possess the same allele at a locus, those alleles also "stack" and present as a single peak. *See*, *e.g.*, Figure 2, below. This is known as allele stacking or allele sharing.

---

authors has noted, "[d]ue to the number of laboratories responding and the federal, state, and local coverage obtained, this MIX13 interlaboratory study can be assumed to provide ***a reasonable representation of current U.S. forensic DNA lab procedures across the community.***" M. Coble, *Interpretation Errors Detected in a NIST Interlaboratory Study on DNA Mixture Interpretation in the U.S.* (2015) ("MIX13 Interpretation Errors") (emphasis in original), available at https://www.nist.gov/document-8.

All participants were provided with the same five mock case scenarios and the same set of five evidentiary DNA profiles to interpret, one for each case. Ground truth was known by the study's authors for each case used in the study. As a result, the study authors were able to assess whether false exclusions or false inclusions were made. Two of the five cases (Case 1 and Case 4) involved two person mixtures, and participants were provided with reference samples for a victim and a suspect who were true contributors to the mixture; for these cases, it was not possible to make a false positive error. For each of the three cases where false positives were possible (Cases 2, 3, and 5)—because known non-contributors were provided among the reference samples—both false inclusions (implicating an innocent person) and false exclusions (excluding the true contributor) were made.

Case 5 involved a four person mixture which, because of significant allele sharing, could be erroneously interpreted as a two person mixture. *Id.* at 29, 30; Figure 2. ***Sixty-nine percent of participants falsely included an innocent individual in this mixture***. *Id.* at 34. If inconclusive opinions are removed from the total, ***92% of participants making a conclusive determination made a false positive error, implicating an innocent individual***.

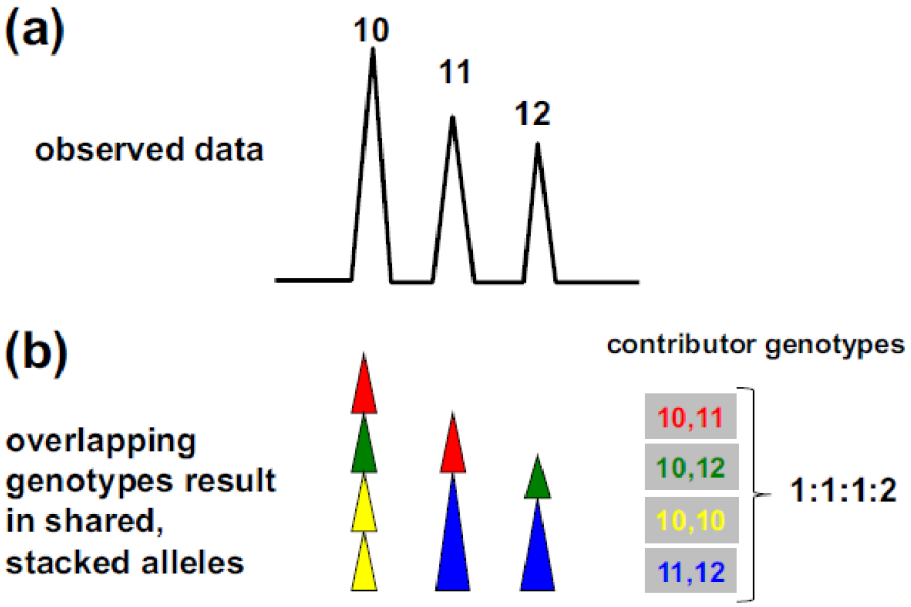

Figure 2. Hypothetical mixture which (a) exhibits only three alleles at a locus (and is thus suggestive of a two person mixture), (b) is actually comprised of three low level contributors plus a single higher level contributor, whose alleles stack on top of one another. Figure from *Interpretation*, at 160, Fig. 7.1.

There are two important consequences of allele sharing. One consequence is that "allele sharing makes accurately deducing the number of contributors to a mixture challenging – and the challenge only grows with each additional contributor to a DNA mixture." *Interpretation,* at 169.[10] Indeed, some scholars have noted that when it comes to mixtures of first-order relatives, the generated profile very easily may *only* look like it originated from fewer contributors than it did. Take this example: "Consider a mixed DNA profile containing a mother, a father, and their biological child. Based on allele count alone and barring mutation, the profile can only ever look like it originated from two individuals, since all of the child's alleles are shared with their parents." M.-H. Lin et al., *The interpretation of mixed DNA profiles from mother, father, child trio*, 44 FORENSIC SCI. INT'L: GENETICS 102175, at 3 (2020), attached as Exhibit CC. While peak heights may provide some hint of an additional contributor, there will often be alternative approximate solutions to account for the peak height variance and reinforce the appearance of two contributors. *Id.* If an analyst cannot accurately determine how many contributors there may be in a mixture, the analyst cannot accurately interpret the mixture. Because the analyst cannot accurately conclude that, for example, a true four-person mixture is not instead a three-person

[10] For example, studies have shown that, because of allelic stacking, more than 75% of known four-person mixtures would be misclassified as two- or three- person mixtures based on the maximum number of alleles detected at any given locus. D. Paoletti et al., *Empirical analysis of the STR profiles resulting from conceptual mixtures*, 50 J. FORENSIC SCI. 1361 (2005). *See also* M. Coble et al., *Uncertainty in the number of contributors in the proposed new CODIS set*, 19 FORENSIC SCI. INT'L: GENETICS 207 (2015), attached as Exhibit X.

mixture, she necessarily cannot deduce the genotype combinations that accurately account for the data under her three-person hypothesis. Put simply, she will always be one person short. Inaccurate interpretation of the mixture impacts not only decisions to include or exclude individuals as potential contributors to the mixture, but also the associated statistical analysis.

And in some situations, the number of contributors' determination itself could be construed as probative: the evidence has different meaning if it is a mixture of four or more family members who lived in the home than if it is a mixture of just the defendant, his wife, and the complainant child. However, a known consequence of first-order relatives making up a mixture is that forensic DNA analysis lacks a reliable method to accurately determine the number of contributors. *See, e.g.*, M. Kruijver & J. Curran, *The number of alleles in DNA mixtures with related contributors*, 61 FORENSIC SCI. INT'L: GENETICS 102748 (2022), attached as Exhibit V.

A second, and related, consequence of high allele sharing amongst first-order relative mixtures is that it becomes increasingly difficult to reliably draw conclusions about the contributors to that mixture. It is known in the forensic DNA community that at some level of allele sharing, forensic DNA analysis stops being able to tell people apart and, instead, falsely identifies known non-contributor relatives as the source of forensic DNA samples. *See, e.g.*, Exhibit CC at 3; *The Kinship Problem*, https://indefenseof.us/issues/kinship-problem (collecting DNA laboratory internal validation studies that have considered the problem of relatedness and STRmix interpretation). Ultimately, and unavoidably, "potential allele sharing from multiple contributors lead[s] to greater uncertainty in the specific genotype combinations that can be reliably assumed." *Interpretation,* at 177. And, as Dr. Butler has unambiguously warned, "[w]hen there is a high degree of interpretation uncertainty from an evidentiary sample, it makes little sense to try and draw conclusions … and expect those conclusions to be reliable." *Id.*

Once the analyst decides to assume a number of contributors and identifies what portions of the DNA data are real, then she formats and submits that data to STRmix for the software to attempt to deconvolute the mixture. STRmix allows an analyst to provide the electronic data to the software without providing any known reference profiles for consideration; in other words, it is possible to ask the software to analyze the data on its own. Doing so will produce a data-driven estimation of the percentage contribution from each contributor and of the best genotype explanations for the data.

IV. FACTUAL BACKGROUND

On April 7, 2021, Jane Doe alleged that a week earlier her father had sexually abused her, as, she further alleged, he had also done in the past. That evening, two Special Agents with the U.S. Army Criminal Investigation Division met with Jane Doe and her mother, Monica Johnston, at their apartment. Exhibit B (DNA Discovery), at JOHNSTON_001857 (filed under seal).[11] The

[11] The documents included in Exhibit B were all produced by the government pursuant to the protective order entered by the Court. The documents contain confidential information about the complaining witness. Accordingly, the defense files these documents under seal, consistent with that protective order. However, because the information about the collection of evidence and DNA analysis described in this section is not itself "sensitive," the parties have agreed that this information need not be redacted from this motion itself.

agents walked through the apartment with Jane Doe, collecting evidence as she described the location and manner of the alleged abuse. *Id.* Based on her allegations, from the living room the agents collected part of a carpet and three couch cushion covers, and from her parents' bed they collected a sheet, three pillowcases, a pillow, and a comforter. *Id*. at JOHNSTON_001834-35. The agents also collected cheek swabs from Jane Doe, Mrs. Johnston, and Mr. Johnston. *Id.*

The agents interviewed Mrs. Johnston. Mrs. Johnston informed the agents that she and Mr. Johnston lived in the apartment with four children: Jane Doe (then age 13, child of Mrs. Johnston with another man); T.L. (then age 12, son of Mrs. Johnston with another man); L.J. (then age 6, daughter of Mr. and Mrs. Johnston); and M.J. (then age 3, son of Mr. and Mrs. Johnston). Mrs. Johnston and Mr. Johnston had sexual intercourse on their bed, and, as Ms. Johnston explicitly informed the agents, all of the children spent time on her and her husband's bed regularly. *Id. at* JOHNSTON_001836, JOHNSTON_001850. Trace evidence analysis also revealed the presence of hairs from the family's dog on every item from the bed (the sheet, pillow, comforter, and three pillowcases). *Id.* at JOHNSTON_000855-57.

The collected items were sent to a U.S. Army Laboratory[12] in Georgia for analysis. From the beginning, the scientific complexities of this case were clear. Even at intake, the Army Lab noted, "[w]hen families are involved, it can complicate our DNA results..." *Id. at* JOHNSTON_001845.

After learning from a case agent that Mr. Johnston was alleged to have sexually abused Jane Doe in the living room and on the bed shared by her parents, U.S. Army Forensic Biologist Sara Green began her analysis of the collected evidence for biological material and eventually DNA.[13] *See id.* at JOHNSTON_001706. Ms. Green assumed, without confirming, that if any semen was present on the comforter, it must be Mr. Johnston's. Thus, her first step was to determine which of the stains contained semen, and her second step was to analyze and interpret only the stains she identified in step one to determine whether any female DNA from Jane Doe was also present.

Ms. Green determined that the couch cushions contained no semen. She determined that the living room carpet sample contained no semen. The flat sheet from the parents' bed, she concluded, contained no semen. She performed no further testing on these items.

Turning to the comforter, Ms. Green first examined it under an alternate light source and determined that 41 stains fluoresced, *i.e.*, might contain biological material. *Id.* at

---

[12] The official title of the laboratory appears to be the "National Defense Forensic Science Center." For ease, and because the biologist who analyzed the DNA here is employed by the Army, we will refer to this laboratory as the "Army Lab."

[13] It also appears from the Army Lab case file that the Biologist had a telephone call with the Assistant U.S. Attorney in July 2021 prior to beginning her analysis. Exhibit B at JOHNSTON_001892. Defense counsel were informed by the Biologist that she only had substantive discussions about the case circumstances with the case agent, not with the prosecutors.

JOHNSTON_001897. Ms. Green did not qualitatively document—either in her notes or with photography—the fluorescence. *Id.* Ms. Green proceeded to conduct serological testing on each of these 41 stains to determine whether there was biological material in the stains that could be consistent with semen.

She began by testing for acid phosphatase, a non-specific screening test that is known to yield significant false positives results. Ms. Green determined that eight of the 41 stains (stains 7, 10, 11, 13, 19, 20, 23, and 25) were positive for AP. *Id.* Again, she did not qualitatively record or photographically document the color change from any of these AP tests. *Id.*

Then, Ms. Green used the Seratec PSA Semiquant Kit to test the eight AP-positive swabs for prostate specific antigen. Ms. Green used the same swabs she had used to test for AP to conduct the test for PSA. She determined that seven of those eight swabs (all but stain 19) were also positive for PSA. *Id.* Again, she did not document the strength of the indicator line, either in her notes or using photography. *Id.*

After the screening tests, Ms. Green advanced the seven PSA-positive stains for additional testing. She prepared slides from the stains, but when she reviewed each of them, she failed to visualize sperm. Exhibit B, at JOHNSTON_001900. Instead of considering the possibility that these results indicated an absence of seminal fluid, Ms. Green reaffirmed her initial position and reported that she had identified semen but had not observed spermatozoa.

Ms. Green then advanced all eight stains that had tested AP-positive for DNA analysis. *Id.* at JOHNSTON_000849. For each stain, a different examiner from the Army Lab performed a two-stage "modified differential extraction" procedure designed to separate epithelial cells from sperm cells. *See* Exhibit F at JOHNSTON_002566-71. Based on the lab's interpretation of each stain's extraction and quantitation results, none of the second fraction testing, which should have contained a higher concentration of male DNA, was advanced to full DNA typing. Each was found to have either no or insufficient DNA for testing. Exhibit B at JOHNSTON_001907. Even though these results undermined Ms. Green's conclusion that she had identified semen stains on the comforter, Ms. Green continued to maintain her categorical serological opinion even in the face of the data. She then carried that unscientific certainty into her DNA analysis.

The interpretation of the differential extraction revealed that none of the eight stains contained large quantities of DNA. Ms. Green ignored the significant uncertainties in each of the samples. She first manually determined that two of the samples were consistent with Mrs. Johnston, the defendant's wife. Five of the remaining stains she called three-person mixtures, and one she called a four-person mixture. These number of contributor calls were unsupported by the data and the case circumstances.

Ms. Green produced a DNA report using STRmix to analyze the six stains that presented as mixtures. *Id.* She did not conduct a standard deconvolution process, in which the analyst inputs the estimated number of contributors for each sample and asks STRmix to analyze the genomes it believes would have combined to create the mixture. Instead, she first assumed for each of the six stains that Mrs. Johnston was a contributor and only ran the probabilities based on that assumption. *Id.* Next, she assumed for each of the six stains that Mrs. Johnston *and* Mr.

Johnston were contributors and ran probabilities on that assumption. *Id.* She never considered whether any untyped first-order relatives might be contributors to any of these mixtures.

According to its expert disclosure, the Government intends to offer Ms. Green's testimony to explain "how a mixture of DNA is detected through testing," to present Ms. Green's "conclusions, based on the testing done in this case, that DNA mixtures were detected in certain samples collected[,]" and to "describe the scientific and professional practices for identifying contributors to DNA mixtures, and the individuals identified as contributors to the DNA mixtures detected in this case." Exhibit A. The disclosure fails to specify the opinions Ms. Green will provide about this case.

Based on her underlying report, we expect that Ms. Green will testify that she concluded that five of the samples she tested from the comforter were mixtures containing DNA from three contributors. Ms. Green is then expected to testify that she provided her assessment of the number of contributors to the STRmix analysis software program and formulated a proposition based on the background information available to her and the validation protocols of her lab, according to which STRmix interpreted the DNA samples. She is expected to testify that STRmix determined, based on her inputs, that given the evidence it was more likely to have originated from Mrs. Johnston, Mr. Johnston, and Jane Doe than from Mrs. Johnston and two unknown, ***unrelated*** people.[14] She is also expected to testify that STRmix determined, based on her inputs, that the DNA in the mixtures was more likely to have originated from Mrs. Johnston, Mr. Johnston, and Jane Doe than from Mrs. Johnston, Mr. Johnston, and one unknown, ***unrelated*** person. Ex. A at 3. There is no indication that Ms. Green evaluated the likelihood that the DNA in these five samples was more likely to have originated from Mrs. Johnston, Mr. Johnston, Jane Doe, and one, two or three additional first-order relatives, or from Mrs. Johnston and one or more unknown ***related*** individuals.

On November 21, 2024, defense counsel met by video with Ms. Green for approximately one hour to discuss her analysis and understand her conclusions. Where facts in this motion are taken from that discussion, it will be explicitly indicated.

The instant motion also makes reference to certain documentation of protocols and validations provided to the defense by the Army Lab. Forensic laboratories are subject to internal validation requirements, *i.e.*, controlled testing involving samples of known DNA mixtures designed to ensure the reliability of laboratory procedures as applied. The initial DNA discovery produced by the government contained no documentation of the Army Lab's protocols or validations. After the defense requested laboratory protocols and internal validation summaries and underlying data for the serological testing conducted, the modified differential extraction

[14] While the government's expert disclosure states that Ms. Green will testify as to the "individuals identified as contributors to the DNA mixtures detected in this case," *see* Exhibit A, we assume that Ms. Green, as a scientist, will not attempt to testify that an "identification" was made, but rather as to various likelihood ratios. Any identification testimony would, of course, be improper.

protocol used, and the version of STRmix used, we received a number of responsive documents (relevant materials are attached as Exhibit F).

The Army Lab's internal materials and Ms. Green's explanation of them on our video discussion contain several significant pieces of information. First, the Army Lab had no protocol instructing analysts how to account for kinship when interpreting DNA samples where multiple family members are potential contributors. *See* Exhibit F, DFSC DNA 114.1, Page 10 ("6.7 Biological relatives – mixtures involving biological relatives that cannot be deduced into individual contributors should be interpreted and reported with caution due to increased allele sharing. Based on the question being asked (*e.g.*, is it the POI versus a relative?), no statistical weight may be available or reportable at the DFSC.").

Second, the Army Lab had not internally validated their STRmix testing protocols to account for a hypothesis that an unknown person who is related to the person of interest contributed DNA to the mixture instead of the person of interest. Defense Meeting with Sara Green, Nov. 21, 2024.

Third, the Army Lab's version of STRmix had not been validated for mixtures of more than four persons. Exhibit F, DFSC DNA 114.1, Page 9 ("6.6.3.2. Mixtures determined to be more than four individuals should not be interpreted further due to mixture complexity.").

Fourth, the serological testing protocols expressly warn about the risk of false positives for semen. Seratec® PSA Semiquant Kit Buffer Validation, Exhibit F, JOHNSTON002589-93 ("It is also well documented that there are reported instances of false positive PSA results in the absence of semen. Sometimes these results are from elevated PSA levels in other bodily fluids due to a biological phenomenon in an individual from a medical condition or using certain products. Other times this is a result of a non-biological substance mimicking a positive result on the test strip.").

Finally, the internal protocols warn about the risk of wrongly estimating the number of contributors in complex samples. *See*, *e.g.*, Exhibit F at DFSC DNA 114.1, Page 9 ("6.6.3 … A mixture… may be determined to be inclusive [if]… [t]he number of contributors cannot be reliably determined.").

V. THE GOVERNMENT'S EXPERT NOTICE FOR MS. GREEN DOES NOT SATISFY RULE 16

On September 10, 2024, the government provided expert notice for Ms. Green. *See* Exhibit A. The expert disclosure fails to provide any of the "opinions" the government intends to elicit from Ms. Green, let alone the specific "bases and reasons for them." This violates the disclosure requirement of Rule 16(a)(1)(G)(iii). On this basis alone, Ms. Green's testimony should be precluded.

In its notice, the government merely lists the *topics* that Ms. Green will testify about. *See* Ex. A ("[Green] is expected to testify generally about the properties of DNA, DNA testing and analysis, and her examination of the DNA recovered in this case…. Green will explain her conclusions, based on the testing done in this case, that DNA mixtures were detected in certain

samples collected. She will describe the scientific and professional practices for identifying contributors to DNA mixtures, and the individuals identified as contributors to the DNA mixtures detected in this case."). It provides no indication what Ms. Green's actual opinions will be as to each of these topics. Such notice was insufficient even before the recent amendment to Rule 16. *See, e.g.*, *United States v. Kaufman,* No. 19 Cr. 504, 2021 U.S. Dist. LEXIS 170367, 2021 WL 4084523, at *19 (S.D.N.Y. Sept. 8, 2021) (quoting *United States v. Valle,* No. 12 Cr. 847, 2013 U.S. Dist. LEXIS 14864, 2013 WL 440687, at *5 (S.D.N.Y. Feb. 2, 2013)).

VI. THE DNA EVIDENCE THE GOVERNMENT SEEKS TO ADMIT SHOULD BE EXCLUDED AS IRRELEVANT AND MISLEADING

Based on our reading of the underlying reports generated by Ms. Green, we anticipate she will testify that it is highly likely that stains on the comforter taken from the complaining witness's parents' bed contain DNA from the complaining witness and her parents. The most fundamental reason to exclude this testimony is that if it is put before the jury, it will make the jury think it means something as to a factual dispute in this case. But it does not.

First, it is irrelevant. What does family members' DNA on an item from their home show? The complaining witness, the defendant, the complainant's mother, and the complainant's three siblings all lived together. All played and rolled and threw clothes and napped on the parental bed. The parents had sexual intercourse on the bed. That DNA of various family members was found on the comforter is consistent with them living in the same home.[15] It does not make the allegation of sexual abuse of the complainant any more or less likely, because the presence of the DNA does not provide any information about how it was deposited there, when it was deposited, whether the various contributors deposited it at different times, or in the course of what activity it was deposited. Indeed, in the forensic context, it is impossible to know even whether DNA was deposited on a given object through direct transfer (*e.g.*, a person touches the object) or indirect transfer (*e.g.*, a dog carries DNA from a person and deposits it on the comforter). *See* Exhibit C at 172-77; *see also* H. Monkman et al., *Presence of Human DNA on Household Dogs and Its Bi-Directional Transfer*, 14 GENES 1486 (2023), attached as Exhibit D. It would be a different analysis if a stain was analyzed that contained sperm and generated a result showing that only the complaining witness and the defendant were contributors. That is

[15] By some estimates, people send roughly 30,000-40,000 of their cells out into the world every minute. *See* ERIN MURPHY, INSIDE THE CELL: THE DARK SIDE OF FORENSIC DNA 21, 21 n. 4 (2015) (citing PATRICIA DANIELS ET AL., BODY: THE COMPLETE HUMAN 46 (2007)). Each of these cells contains DNA that can transfer "from one surface or person to another, and this can potentially happen several times." Butler, Exhibit BB at 9. Moreover, "[i]mprovements in DNA testing methods have allowed forensic scientists to reduce the quantity of DNA required for profiling an individual." *Id*. at 2. A DNA profile can now be obtained from just a few skin cells. *Id.* Such sensitivity makes it more likely that analysts will "detect small amounts of DNA, including background DNA that may be present in the environment." *Id.* at 8. *See Walker v. City of New York*, No. 14-cv-680-NRM, ECF No. 255, 2024 WL 4198451 (E.D.N.Y. Sep. 16, 2024), at *21-23.

not the case here. Even accepting the reliability of the DNA analysis—which we do not—what it shows is that Mr. Johnston, his wife, and their daughter were all on the bed at some point. It does not show that Mr. Johnston and the complaining witness were on the bed alone at any point, nor does it show that they engaged in sexual activity. There is no allegation that Mrs. Johnston was ever involved in the sexual abuse, so a mixture of her DNA with her husband's and daughter's does not support the government's theory. This evidence thus fails to satisfy Federal Rule of Evidence 401.

Equally problematic is this evidence's capacity to mislead. An expert witness flying in from Georgia and taking the jury through the multi-step analysis she engaged in using various tests and probabilistic modeling will make the jury assume, understandably, that this DNA evidence actually means something relevant to these charges, and in particular that the likelihood ratios produced by STRmix prove something at issue in this case. Because they do not, and because, for all the reasons set forth below, this DNA analysis is unreliable, it is more prejudicial than probative, and it must be excluded under Rule 403.

VII. THE DNA EVIDENCE THE GOVERNMENT SEEKS TO ADMIT IS INSUFFICIENTLY RELIABLE TO SATISFY RULE 702 AND *DAUBERT*.

A. The Army analyst did not take kinship into account when interpreting the DNA mixtures because the Army Lab had not internally validated a procedure for doing so, rendering her conclusions unreliable.

What jumps out immediately about the DNA analysis attempted here is the complexity of having so many first-order relatives as potential contributors. Because the Army Laboratory has never internally validated protocols for interpreting DNA potentially contributed by two parents and children of varying biological relation, Ms. Green's conclusions about the mixtures on the stains on the comforter are inherently unreliable. This is true in particular because the lab's lack of internal validation prevented her from substantively grappling with the relatedness impact on her number of contributor analysis and from testing defense hypotheses that considered the possibility that an unknown relative of the person of interest contributed the DNA in question, rather than the person of interest.

The general procedures of the Army Lab for interpreting DNA evidence viewed in "Globalfiler" (DFSC DNA 114.1, attached in Exhibit F) contain only one paragraph relating to kinship, which basically warns the analyst to proceed with caution where biological relatives are potential contributors, without providing detailed guidance or a rationale.

> **6.7 Biological relatives** - mixtures involving biological relatives that cannot be deduced into individual contributors should be interpreted and reported with caution due to increased allele sharing. Based on the question being asked (e.g., is it the POI versus a relative?), no statistical weight may be available or reportable at the DFSC. Consult with the Technical Leader where needed.

Exhibit F, DFSC DNA 114.1, Page 10.

Rather than proceeding with caution, Ms. Green failed to pursue the most basic information she had that bore directly on the number of contributors assessment: information about who lived in the household. As the NIST *Human Factors* report emphasizes, "fact patterns that increase the relatedness risk should trigger additional information gathering." Exhibit C at 65. "When DNA is recovered from bedding or clothing in a familial sexual assault case, relevant questions include: How many first-order relatives (both adults and children) live or have regular access to the home? Are the items commingled in the laundry along with those of other family members? Can reference samples from the family members be obtained?" *Id.* Where elimination samples are not available, as here, the analyst should seek additional case information about who could be an alternate source of the DNA and should carefully consider the resulting propositions. If meaningful to the case, the defense hypothesis should involve a relative and not just an unknown, unrelated individual. *See* T. Kalafut et al., *Investigation into the effect on mixtures comprising related people on non-donor likelihood ratios and potential practises to mitigate providing misleading opinions*, 59 FORENSIC SCI. INT'L: GENETICS 102691 (2022), attached as Exhibit G.

As explained in the recent *Human Factors* report,

> Propositions are key to evaluating findings in accordance with the principles of interpretation. Once the analyst has defined the relevant issue(s) based on the case circumstances, they can start to formulate propositions to assess the findings. Propositions and proposition pairs should have the following properties:
>
> 1. They should be based on the available case information.
> 2. They should be mutually exclusive (*i.e.*, they cannot both be true simultaneously).
> 3. Although not necessarily exhaustive, they should cover all reasonable options dictated by the case circumstances and any other relevant known background. This means that multiple sets of propositions may be needed to properly evaluate the findings.
> 4. They should be explicit and leave no doubt about what is being proposed.
> 5. They should be structured so as not to transpose the conditional.

Exhibit C at 62 (internal citations omitted).

The report specifically advises that when, as here, elimination samples are not available, the analyst should carefully consider alternative meaningful propositions that include relatives of different degrees. *Id. at* 65–66**.** A series of published studies underscores the point that STRmix cannot reliably interpret DNA mixtures when analysts fail to propose reasonable defense hypotheses. *See*, *e.g.*, J. Buckleton et al., *Helping formulate propositions in forensic DNA analysis*, 54 SCI. & JUST. 258, 259-60 (2014) ("[P]ropositions emanate from the case information given by both parties…. [T]he forensic scientist should select a reasonable proposition consistent with defence's view (and case information). If this proves to be a poor choice subsequently, the findings should be re-assessed."), attached as Exhibit I; J. Buckleton et al., *Guiding proposition setting in forensic DNA interpretation*, 62 SCI. & JUST. 540, 546 (2022) ("[T]he use of a

conditioning profile…greatly improves the ability to discriminate between true and false donors."), attached as Exhibit J; J. Buckleton et al., *When evaluating DNA evidence within a likelihood ratio framework, should the propositions be exhaustive?,* 50 FORENSIC SCI. INT'L: GENETICS 102406 (2021) (propositions should account for the presence of an assumed contributor when there is a "reasonable prior probability" that their DNA is in the mixture, a relatively low standard), attached as Exhibit K; P. Gill et al., *DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence – Guidelines highlighting the importance of propositions Part I: evaluation of DNA profiling comparisons given (sub-) source propositions*, 36 FORENSIC SCI. INT'L: GENETICS 189 (2018), attached as Exhibit L.

Ms. Green did not use any hypotheses in STRmix that accounted for the possibility that a relative of the persons of interest contributed DNA to the samples. Instead, she proposed only two hypotheses to STRmix. Under her first hypothesis, Ms. Green conditioned each sample on Mrs. Johnston's profile, *i.e.*, she assumed that Mrs. Johnston was a contributor to every sample. Ms. Green never evaluated any of the samples under the hypothesis that Mrs. Johnston was *not* present in each sample. Thus, for each sample, she asked the software to consider the likelihood of obtaining the evidence if the sample contained a mixture of Mrs. Johnston, Mr. Johnston, and Jane Doe versus if the sample contained a mixture of Mrs. Johnston and two unknown, *unrelated* individuals. Under her second hypothesis, she extended her assumptions and conditioned each sample on Mr. and Mrs. Johnston's combined presence in every sample, *i.e.*, she assumed that both Mr. and Mrs. Johnston had contributed to every sample. Thus, she then asked the software to consider the likelihood of obtaining the evidence if the sample contained a mixture of Mrs. Johnston, Mr. Johnston, and Jane Doe versus if the sample contained a mixture of Mrs. Johnston, Mr. Johnston, and one unknown, unrelated individual.

Perhaps these choices would be understandable for an analyst who had been provided no background information about potential contributors to these samples. After all, the mechanisms of DNA shedding and transfer are still poorly understood, such that a detectable amount of DNA from a one-time visitor to the house (perhaps a friend of one of the children, a maintenance worker, or a grandparent) could foreseeably end up on the comforter even if the visitor never touched the comforter. *See* ERIN MURPHY, INSIDE THE CELL: THE DARK SIDE OF FORENSIC DNA 21, 21 n.4 (2015) (citation omitted); M. Zacher et al., *Transfer and persistence of intruder DNA within an office after reuse by owner*, 73 FORENSIC SCI. INT'L: GENETICS 103130 (2024), attached as Exhibit M; B. Szkuta et al., *The presence of background DNA on common entry points into homes*, 7 FORENSIC SCI. INT'L SUPP. 784, 784-86 (2019), attached as Exhibit N. But before Ms. Green performed any testing in this case, she knew not only the identities of and family relationships between the members of the household; she also specifically knew that there was reason to believe that DNA from each of the four children could reasonably be present on any samples taken from items on the bed. Based on the case scenario information Ms. Green had available, and even assuming her number of contributors' assessment was reliable and accurate (which it was not), it would have been far more reasonable for Ms. Green to more deeply investigate her likelihood ratio propositions. For example, Ms. Green never deconvoluted any of the samples *without* conditioning on Mrs. Johnston. While conditioning is advisable when it can

be reasonably assumed, this specific scenario complicated Ms. Green's conditioning assumptions. There were four potential contributors in this environment who are first-order relatives of Mrs. Johnston. Ms. Green's failure to examine the mixture without conditioning meant that she unreliably constrained her interpretation and could not evaluate the additive impact of alternative hypothesis. *See*, *e.g.*, Sacramento County District Attorney's Crime Laboratory, Internal Validation of STRmix V2.4 (2017), available at https://indefenseof.us/uploads/Sacramento-Cty-DA-STRmix-V2-4-internal-validation-summary.pdf (explaining that their internal validation examining high allele sharing and the impact of relatedness revealed that the failure of the combined LR to be additive may be a diagnostic for the impact of high allele sharing).

Beyond her failure to investigate alternative hypotheses, the combined likelihood ratio, and addictiveness, Ms. Green also failed to propose any of the following reasonable defense hypotheses, assuming that the prosecution hypothesis remains the same:

1. The sample contains a mixture of three or more related individuals.
2. The sample contains a mixture of Mrs. Johnston and two unknown **related** individuals.
3. The sample contains a mixture of Mrs. Johnston, Jane Doe, and one unknown **related** individual.
4. The sample contains a mixture of Mrs. Johnston, Mr. Johnston, and one unknown **related** individual.
5. The sample contains a mixture of Mrs. Johnston, Jane Doe, and one unknown **unrelated** individual.
6. The sample contains a mixture of Mrs. Johnston, Mr. Johnston, and T.L.
7. The sample contains a mixture of Mrs. Johnston, T.L., and one unknown **related** individual.
8. The sample contains a mixture of Mrs. Johnston, T.L., and one unknown **unrelated** individual.
9. The sample contains a mixture of Mrs. Johnston, L.J., and one unknown **related** individual.
10. The sample contains a mixture of Mrs. Johnston, L.J., and one unknown **unrelated** individual.
11. The sample contains a mixture of Mrs. Johnston, M.J., and one unknown **related** individual.
12. The sample contains a mixture of Mrs. Johnston, M.J., and one unknown **unrelated** individual.
13. The sample contains a mixture of Mrs. Johnston, L.J., and Jane Doe.
14. The sample contains a mixture of Mrs. Johnston, M.J., and Jane Doe.
15. The sample contains a mixture of Mrs. Johnston, T.L., and Jane Doe.

Each of these hypotheses is exonerating because it proposes an explanation for the evidence that is inconsistent with the alleged offense conduct. Ms. Green's failure to consider these propositions—especially those not requiring unavailable conditioning profiles—creates a significant risk that the reported likelihood ratios will mislead jurors about the strength of the

evidence obtained from the comforter, even to the point of providing extremely strong evidence for a false inclusion of Mr. Johnston.

The danger posed by Ms. Green's failure to investigate the impact of relatedness is more significant than a mere threat of slightly inflated likelihood ratio reporting. In fact, empirical data has shown that a failure to account for the potential presence of first-order relatives in a DNA mixture can create highly misleading false inclusionary likelihood ratios. *See*, *e.g.*, Palm Beach County Sheriff's Office Laboratory, Internal Validation of STRmix™ V2.4 (2017), attached as Exhibit O; Jefferson County Regional Crime Laboratory, Internal Validation of STRmix™ V2.6 for the analysis of GlobalFiler™ profiles, attached as Exhibit P; Las Vegas Metropolitan Police Department, Internal Validation of STRmix™ v2.6 (QIAGEN Investigator® 24plex QS with 3500xl) (2020), attached as Exhibit Q. Moreover, there is no empirical data demonstrating that, where conditioning profiles for all potential contributors are not available, STRmix can reliably distinguish between true contributors and true noncontributors when the sample includes DNA from multiple first-order relatives. In fact, the available empirical data suggests that STRMix cannot reliably distinguish among first-order relatives even when conditioning profiles are available. *See*, *e.g.*, Exhibit O (demonstrating that even when conditioning, it is possible to obtain high false-positive likelihood ratios for known noncontributor relatives); Exhibit P at 29 ("[E]ven when conditioning information is used false exclusion of a true contributor and the high LRs in favour of inclusion for some non-contributors persist.").

When counsel asked Ms. Green during the November 21, 2024 conference call about her failure to take steps to account for the presence of first-order relatives in her interpretation, her answer was simple: the Army Lab has not validated its STRmix testing protocol to account for relatedness. Other labs have conducted that validation; the Army Lab has not.

Every lab that has empirically described the impact of relatedness has discovered a significant risk of false positives that will not be flagged by any known diagnostic. *See* The Kinship Problem, https://indefenseof.us/issues/kinship-problem. In fact, the Las Vegas Metropolitan Police Department's lab responded to its internal validation data by requiring the following disclaimer to be included in its reports when "reporting samples which may contain related individuals:" "In DNA mixtures of closely-related individuals (such as parents, offspring, and siblings), false positive inclusions of other closely-related family members can occur due to the elevated sharing of genetic information between relatives." Las Vegas Metropolitan Police Department Forensic Laboratory, Biology/DNA Detail Procedures Manual (2024), attached as Exhibit T, at 272. Because Ms. Green had no reference profiles for the other three children and the Army Lab failed to consider relatedness in their internal validation, Ms. Green did not factor in the impact of high allele sharing on her analysis or the possibility that those three children could have contributed DNA to the mixtures.

In this case, the degree of allele sharing—even among the three samples that have been collected—is exceedingly high. Jane Doe shares 69% of her DNA with the combination of her mother and the defendant. *See* Exhibit B, 21-0724 STRMix_references (containing the typed profiles of Jane Doe, the defendant, and the defendant's wife/Jane Doe's mother and reflecting that Jane Doe shares 29 out of 42 typed alleles with either her mother or Mr. Johnston or both).

The defendant shares 52% of his DNA with the combination of Jane Doe and her mother, despite not being biologically related to them. *See id.*, 21-0724 STRMix_references (containing the typed profiles of the defendant, Jane Doe, and the defendant's wife/Jane Doe's mother and reflecting that the defendant shares 22 out of 42 typed alleles with either his wife or his step-daughter or both). And the defendant's wife/Jane Doe's mother shares a whopping 80% of her DNA with the combination of the defendant and the complainant child. *See id.*, 21-0724 STRMix_references (containing the typed profiles of the defendant's wife/Jane Doe's mother, the defendant, and Jane Doe and reflecting that the defendant's wife/Jane Doe's mother shares 34 out of 42 typed alleles with either her husband or her daughter or both).

Where such high degrees of allele sharing are present and so many first-order relatives are in the environment, the literature suggests that forensic testing may have reached the edges of its limitations. *See* Exhibit CC at 3 ("There are limits of DNA typing and allele sharing with relatives. They are inherent to the use of DNA, and are caused by well-known effects of genetics. Interpretation, by whatever means, can only work within these limitations."). Certainly, where no internal validation studies have been conducted by the lab to characterize the limits of the forensic method, the analyst must acknowledge that "[a]ll scientific methods have limits." *See* Exhibit BB at 2.

Before conclusions grounded in a particular methodology can be admitted into evidence, the methodology must be supported by "appropriate validation." *Daubert,* 509 U.S. at 590. Because the laboratory analyzing these samples had no validated protocols for taking account of the specific case situation presented and because Ms. Green failed to take account of alternative hypotheses in analyzing the DNA, the government cannot meet its burden of showing that the resulting analysis is reliable. The DNA analysis must be excluded.

B. <u>The analyst's conclusion that there were three contributors to certain stains on the comforter is unreliable and renders the downstream conclusions unreliable.</u>

In addition to complicating deconvolution of complex mixtures, high allele sharing makes the reliable determination of the number of contributors impossible. In this case, Ms. Green's determination of the number of contributors could itself be construed as probative: the evidence has a different meaning if it is a mixture of four or more family members who lived in the home than if it is a mixture of just the defendant, his wife, and the complainant. Because it was not reached pursuant to scientifically reliable methods, Ms. Green's number of contributor determination was unreliable.

Here, not only is the assessment of the number of contributors to each mixture on the comforter potentially deceptive, it also influenced every conclusion that followed. Because it was unreliable, it renders Ms. Green's conclusions unreliable.

1. The Number of Contributors Decision Is Always Subjective

The true number of contributors ("NOC") to a forensic sample can never be known with certainty absent knowledge of what actually happened, as the Army Lab's own protocols make clear. *See* Exhibit F, DFSC DNA 114.1, Page 7 ("6.3 Determine the minimum number of

contributors (n) *where possible*") (emphasis supplied). Analysts must estimate the number of contributors in a sample based on observed characteristics of the DNA profile. These characteristics include the maximum number of alleles at any given locus; peak height ratios; the potential for artifacts such as allelic stutter, drop-in, or drop-out; and the estimated frequencies of alleles within the relevant population. *Id.* 7-8. This analysis requires a number of subjective decisions by the analyst about what to include. "[A]n error in determining how many individuals' DNA were present within a mixture could have rippling effects through subsequent stages of analysis, affecting decisions [by the algorithm] about differentiating signal from noise and distinguishing which DNA came from which individual." Exhibit E, *Kwong*, supra, at 291.

2. *Because Sexual Abuse Was Alleged, Ms. Green Wrongly Assumed That Positives On The Serological Testing Indicated That Semen Must Be Present, Skewing Her Analysis Toward The Presence Of Mr. Johnston.*

"Case information can influence the estimated NOC[,] … DNA profiles of assumed contributors, analyst experience, and additional information such as the appearance of peaks visible below the [analytical threshold]. These factors tend to make manual NOC estimation highly variable. With respect to case information and assumed contributors, the analyst should clearly document the case information that was known and considered during the interpretation steps." Exhibit C, *Human Factors*, at 50 (citing J. Butler et al., *Interlaboratory Studies Involving DNA Mixtures (MIX05 and MIX13): Variation Observed and Lessons Learned*, 37 FORENSIC SCI. INT'L: GENETICS 81, 81–94 (2018), attached as Exhibit R). Here, Ms. Green's number of contributors assessment was influenced by her belief that semen was present on stains on the comforter, leading her to conclude that Mr. Johnston had to be a contributor. This belief was faulty.

Before beginning the DNA testing, Ms. Green performed two screening serological tests, one for AP and one for p30 antigen. Both of these substances occur naturally in semen, but both can be detected when present because of a non-semen source. *See e.g.*, P. Hooft & H. van de Voorde, *Interference of body products, food and products from daily life with the modified zinc test and the acid phosphatase test*, 66 FORENSIC SCI. INT'L 187, 188–95 (1994) (AP tests react positively with each of the following non-semen materials: whole blood; blood plasma; blood serum; blood-contaminated feces; uncontaminated vaginal discharge; vaginal discharge containing menstrual blood; cauliflower; chicory; leek; salad; savoy; spinach; sprouts; beef; goat meat; horse meat; pork; mutton; Boursin cheese; cacao/chocolates; cinnamon; nutmeg; parsley; polluted canal/river water; and water from a public swimming pool.), attached as Exhibit H; A. Stroud et al., *A comprehensive study into false positive rates for 'other' biological samples using common presumptive testing methods*, 63 SCI. & JUST. 414, 417 (2023) ("All samples of faeces gave a positive reaction in the AP test…. All male urine (8 samples) tested gave a weak positive result for AP.… Two female participants (6 urine samples) gave weak positive result for AP.… 55% of vaginal material samples gave a positive result … for AP."), attached as Exhibit S; K. Ostapovic, *Cross-Reactivity with Food Products, Fungal Growth, and Acid Phosphatase Testing* (2021) (M.S. thesis, Boston University School of Medicine) (on file with Boston University Libraries) ("Numerous types of fungi and other foods were shown to have detectable levels of

[AP] in the absence of semen. Evidence that has been stored improperly and has developed mold, or evidence that has come into contact with mold spores or food items at a crime scene… should be viewed with a conservative mindset when positive AP results are obtained."); Seratec® PSA Semiquant Kit Buffer Validation, attached as Exhibit F, at JOHNSTON002589-93 ("It is also well documented that there are reported instances of false positive PSA results in the absence of semen. Sometimes these results are from elevated PSA levels in other bodily fluids due to a biological phenomenon in an individual from a medical condition or using certain products. Other times this is a result of a non-biological substance mimicking a positive result on the test strip…. The literature also recommends not testing an item that has come into contact with "forensic detection aids" like acid phosphatase (AP). In case work a false positive result can result in the false identification of semen."); SERATEC®, *PSA In Body Fluids*, available at https://www.seratec.com/docs/user_instructions/psa_in_body_fluids (manufacturer's summary of peer-reviewed research confirming that p30 can be detected in vaginal fluid, saliva, urine, blood serum, amniotic fluid, breast milk, and breast secretions), attached as Exhibit U.

Ms. Green determined that swabs of eight stains on the comforter turned purple when exposed to the AP testing reagent. Exhibit B at JOHNSTON_001897. She did not document whether the AP fluorescence was weak, strong, or somewhere in between. *Id.* She then used the same swabs to test for the presence of p30; seven of the eight stains were positive on the p30 test. *Id.* After these two screening tests for semen, Ms. Green performed two additional tests. First, she visualized the samples under a microscope in an effort to detect sperm cells. In each sample, she visualized zero sperm cells. Ex. B at JOHNSTON_001897. Then, she had the lab perform a two-stage differential extraction on each sample to attempt to separate epithelial cells from any sperm cells. Each time, the testing produced little to no DNA in the "male fraction" of the stains during the differential extraction. *Id.* at JOHNSTON_001899-1907.

During the conference with defense counsel on November 21, 2024, Ms. Green informed counsel that she assumed that all the stains that were positive for both AP and p30 contained semen. She made—and stuck to—this assumption despite considerable evidence of false positive results generated by both of the tests she used, as recognized by the internal validation studies of her own laboratory. *See* Exhibit F at JOHNSTON_002589, 002592-93 ("It is also well documented that there are reported instances of false positive [p30] results in the absence of semen."). She maintained this assumption even though she did not visualize any sperm under a microscope from any of these stains and even though she did not recover significant quantities of male DNA in her later DNA typing. Ms. Green also maintained her assumption in the face of a stain (13) that tested positive on both serological screeners but was analyzed as containing DNA from a single source: Monica Johnston. Exhibit B, at JOHNSTON_001696, 001897, 001912. As she explained to counsel during the November 21, 2024 conference call, because she assumed the positive tests indicated semen, Ms. Green further assumed that any semen on the comforter must belong to Mr. Johnston. She reached this conclusion despite the known presence of at least two other biological males in the home, one of whom was old enough for puberty to have begun. Thus, before she began interpreting any of the DNA recovered from these samples, she expected to find Mr. Johnston's DNA in each of them, because she expected his semen to be present in each.

But because Ms. Green detected no sperm on either additional test she performed, and because a sample could return a false positive on each screening test used here for well-documented reasons unrelated to the presence of semen, Ms. Green's assumption that this substance must be semen remained merely an assumption to be tested, not one that was confirmed. Further, just because "a sample contains biological material or cell types does not necessarily mean that the DNA profile must have been derived from that material." Exhibit C at 111. This flaw rendered the rest of her analysis unreliable. Moreover, her failure to test other hypotheses as to the source of any potential semen renders her secondary assumption (that Mr. Johnston must have deposited any semen present on the comforter and thus should be assumed as a contributor to any sample containing semen) similarly unreliable.

3. *Allele Sharing—Especially Due To Kinship—Significantly Complicates A Reliable Estimate of The Number of Contributors.*

Ms. Green then used her assumption that semen was present and that it must be from Mr. Johnston to drive her number of contributors' determinations. As discussed above, one of the biggest complications when it comes to estimating the number of contributors in any sample is allele sharing. *Interpretation* 153. It is well-documented that the presence of related individuals exacerbates the risk of misunderstanding the maximum allele count and the number of contributors. *See* Exhibit V.

The allele sharing problem comes into sharp relief in a case scenario like this one, where two of the household members are the full biological children of two other household members. L.J. and M.J. share 100% of their alleles with Mr. and Mrs. Johnston. This means, for example, that if DNA from L.J. and DNA from M.J. were present in a hypothetical four-person mixture with DNA from Mrs. Johnston and DNA from Mr. Johnston, then at each locus of analysis, all of L.J.'s alleles and all of M.J.'s alleles would "stack" on top of the identical alleles from one or both of their parents. In this situation, unless an analyst already knew that the mixture contained DNA from two parents and two of their biological children, one might expect the analyst to estimate that the mixture contained DNA from just two people (Mr. and Mrs. Johnston), because two of the people in the mixture (the children) share 100% of their alleles with the other two people in the mixture. *See* Exhibit G.

In this situation, high allele sharing makes an accurate assessment of the number of contributors impossible to propose. The analyst has to look at peak heights on the EPG to see if there is a significant imbalance, but this too introduces a high level of subjectivity. *See* J-A. Bright et al., *Internal validation of STRmix – A multi laboratory response to PCAST*, 34 FORENSIC SCI. INT'L: GENETICS 11 (2018) (31 of 31 laboratories underestimated the number of contributors to a known six-person DNA mixture), attached as Exhibit W. But with first-order relatives, peak heights will also predictably fail the analyst. *See*, *e.g.*, Exhibit CC at 3. Because of the level of uncertainty introduced by first-order relatives, and the failure to take the other potential relatives into consideration, Ms. Green's assumption about the number of contributors cannot be shown to be reliable.

4. Regardless of Kinship, Estimating the Number of Contributors to Complex Mixtures Becomes More Difficult as the Complexity of the Mixture Increases.

Even outside the context of first-order relatives, an analyst's ability to estimate accurately the number of contributors to a mixture diminishes as the complexity of the mixture increases. As relevant here, research on complex DNA mixtures has demonstrated that trained and qualified forensic scientists very often underestimate the number of contributors in known six-person samples. Four recent peer-reviewed studies support this conclusion. *See* Exhibit W (reporting that 100% of participating laboratories (31 of 31) misidentified a known six-person mixture: 28% interpreted the mixture as having five contributors, 69% as having four contributors, and 3% as having three contributors); Exhibit X (finding that between 86% and 94% of mixtures known to contain six contributors were misidentified by analysts as containing five or fewer contributors); G. Dembinski et al., *Estimating the number of contributors of theoretical mixture profiles based on allele counting: Does increasing the number of loci increase success rates of estimates?*, 33 FORENSIC SCI. INT'L: GENETICS 24 (2018) (finding that known five-person mixtures were analyzed as having fewer than five contributors 43% of the time, and that known six-person mixtures were analyzed as having fewer than six contributors 92% of the time), attached as Exhibit Y; L. Brinkac et al., *DNAmix 2021: Laboratory policies, procedures, and casework scenarios summary and dataset*, 48 DATA IN BRIEF 109150 (2023) (finding that 97% of participating laboratories (83 of 86) instruct analysts to assess number of contributors manually, and observing considerable discrepancies in laboratories' standard operating procedures for assessing the number of contributors for complex samples), attached as Exhibit Z. *See also* R. Hicklin et al., *Variation in assessments of suitability and number of contributors for DNA mixtures*, 65 FORENSIC SCI. INT'L: GENETICS 102892 (2023) ("For the five- and six-person mixtures in this study, we can conclude that assessing [the estimated number of contributors] precisely is not a reasonable expectation…."), attached as Exhibit AA.

Here, six individuals had regular contact with the comforter. The studies referenced above demonstrate that even if these six individuals were all completely unrelated to each other, a scientist would very likely underestimate the number of contributors to a mixture containing DNA from all six of them. But the scenario here is far more complicated than that because of the number of first-order biological relationships between the people with regular access to the comforter. The analysis here was complicated even further by the fact that the analyst assumed, on the basis of screening tests, that semen from one of the biological parents was present in the samples, which drove her failure to account for alternative possibilities. Accordingly, since it is a virtual certainty that a scientist would underestimate the number of contributors to a mixture containing DNA from six related contributors, the government cannot establish that Ms. Green's number of contributors' assessment in this case is reliable.

5. When the Number of Contributors Is Underestimated, as Likely Happened Here Because of the Presence of First-Order Relatives, the Ensuing DNA Analysis Is Unreliable.

If an analyst cannot accurately determine how many contributors there may be in a mixture, the analyst cannot accurately interpret the mixture. For example, if an analyst cannot

accurately conclude that certain samples contain DNA from three contributors instead of four, five or six contributors, the software into which she inputs this data necessarily cannot deduce the genotype combinations that accurately account for the data. Put simply, this is because the software will be trying to sort the data into three contributor piles when instead it should be trying to sort the data into four, five or six contributor piles. Inaccurate assessment of the number of contributors impacts not only decisions to include or exclude individuals as potential contributors to the mixture, but also the associated statistical analysis. "[A]llele sharing from multiple contributors lead[s] to greater uncertainty in the specific genotype combinations that can be reliably assumed." Butler, *Interpretation*, at 177. The uncertainty rises significantly with each new contributor to the sample. *See* Exhibit BB at 31. And "[w]hen there is a high degree of interpretation uncertainty from an evidentiary sample, it makes little sense to try and draw conclusions . . . and expect those conclusions to be reliable." *Interpretation* 177.

The most relevant study to this case is appended hereto as Exhibit G. The authors there found that analysts often underestimate the number of contributors to samples including first-order relatives. The authors observed that when "the relatives of the [person of interest] are in fact true donors to the mixture, there is an increased probability of high adventitious support for the [person of interest,]" and that this high adventitious support is most likely to occur when the profile of a non-donor person of interest is compared to a mixture of DNA donated by multiple relatives of the non-donor person of interest. *Id.* at 2. In other words, a false positive inclusion is most likely to occur when the true donors to a DNA mixture are first-order relatives of the non-donating person of interest. Accordingly, the government cannot establish that the interpretation of this mixture was reliable.

6. The Government Cannot Establish That the Number of Contributors' Assessment In This Case Is Reliable.

Because there is no objective way to determine whether a given peak on an electropherogram comes from one or multiple contributors, and because of the kinship problem, the government cannot establish that Ms. Green reliably determined the number of contributors to samples that may have been contributed to by multiple first-order relatives. Ms. Green's inability to reliably apply that methodology to this case is readily apparent in her analysis of the number of contributors to stain 10, which depended at least in part on her assumption about the identity of one of the contributors—Mr. Johnston.

Stain 10 is one of the six stains on the comforter that Ms. Green advanced to DNA interpretation. The documentation shows that when Ms. Green first interpreted stain 10 using STRmix, STRmix found a likelihood ratio of zero for Mr. Johnston as a contributor when it was assumed that Mrs. Johnston was a contributor. Exhibit B, SG9053.1_21SG0724Q5-10F1_Comforter_stain_10F1_DEH_3500B_2021-08-10_GF_C01.csv_Results-1. But this conflicted with Ms. Green's assumption that Mr. Johnston must be included as a contributor to the stain, which was based on her threshold assumption that the stain contained semen (which in turn was based on the unreliable serological tests).

So, Ms. Green looked at each of the per locus likelihood ratios to figure out why STRmix was telling her that Mr. Johnston could not be included as a contributor to stain 10. She observed that the alleles appearing at D1S1656 in the evidentiary sample did not match Mr. Johnston's genetic profile because there was no 14.3 allele found at that locus in the evidentiary sample. In other words, if Mr. Johnston's DNA had been present in the evidentiary sample, there would have been a 14.3 allele at that locus. Because this zero likelihood ratio conflicted with her assumption that Mr. Johnston must be a contributor to any stains testing positive for both AP and p30, Ms. Green marked this STRmix analysis as "failed." *Id.* Ms. Green then adjusted the analysis so that it would match Mr. Johnston's profile. She instructed STRmix to ignore the complicating locus (D1S1656). Exhibit B, SG9053.1_21SG0724Q5-10F1_Comforter_stain_10F1_DEH_3500B_2021-08-10_GF_C01.csv_Results-2; JOHNSTON_1910.

In the conference with defense counsel on November 21, 2024, Ms. Green informed counsel that instructing STRmix to ignore a locus, as she did here, is an option that should be used very rarely. She explained that based on her review of the data, Mr. Johnston's alleles appeared at many other loci in the evidentiary sample, leading her to believe that the missing 14.3 at D1S1656 was merely a copying error that sometimes happens during the testing process. However, because of Ms. Green's assumption that all the stains testing positive for both AP and p30 *must* contain semen, she altered the data to include Mr. Johnston at locus D1S1656 instead of accounting for the possibility that the missing 14.3 allele could be explained by the fact that a relative or relatives of Mr. Johnston deposited the alleles she observed at the other loci that she assumed were from Mr. Johnston. Moreover, Ms. Green did not re-run the analysis with the filter off, nor did she re-extract the stain, nor did she consider or ask STRmix to consider that the contributor was actually a biological child of Mr. Johnston, who would have similar but not identical alleles. And when Ms. Green sent her analysis for peer review, she included a note to the peer reviewer to ignore the locus that excluded Mr. Johnston. Exhibit B at JOHNSTON_001926. This method of analysis – rejecting any information that contradicts one's working theory, rather than considering other viable hypotheses based on known information – is inherently unreliable. Accordingly, Ms. Green's testimony should be excluded.

C. STRmix has not been validated for interpreting DNA samples involving more than five people, let alone more than five relatives.

Because the number of contributors' assessment here is inherently unreliable, the government cannot reliably establish that the DNA samples in this case contained fewer than five contributors. *See supra* Part VII.B. Thus, in these circumstances, it is impossible for an analyst to determine with any degree of scientific certainty that the samples from the comforter contain DNA from fewer than five people. The government therefore cannot sustain its burden in showing that the Army Lab's analysis in this case is reliable, because STRmix has never been generally validated for mixtures of six or more persons (*see United States v. Ortiz*, __ F.Supp.3d __, 2024 WL 2889873, at *22-30 (S.D. Cal. June 9, 2024)), and separately because the lab's STRmix protocol had not been internally validated for mixtures of five or more persons. *See* Exhibit F, DFSC DNA 114.1, Pages 6-10.

The problematic nature of determining accurately the number of contributors when multiple first-order relatives are potential contributors is especially significant to this case given the tendency of STRmix, at the interpretation stage, to provide high adventitious support for the person of interest when a relative is the true contributor. *See* Exhibit G.

A lab must validate a probabilistic genotyping program on mixtures containing a given likely number of contributors before the system can be applied to such mixtures. *See* Scientific Working Group on DNA Analysis Methods ("SWGDAM") Guidelines for Validation of Probabilistic Genotyping Systems § 4.1.6.3 ("The number of contributors evaluated should be based on the laboratory's intended use of the software. A range of contributor numbers should be evaluated in order to define the limitations of the software."). This is because the difficulty of reliably interpreting samples increases when the number of contributors increases, or when the proportion of the sample attributable to a minor contributor decreases. PCAST, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-STR Comparison Methods* 80-81 (2016) (noting that scientific validity has only been established within the specific range for which experimental evidence of validity is available); *People v. Hillary*, Ind. No. 2015-15, slip op. at 8 (St. Lawrence Cty. Ct., N.Y., Aug. 26, 2016) (holding that "the lack of internal validation by the New York State Police crime lab . . . precludes the use of the STRmix results"). And even "a laboratory's satisfaction with its validation results does not" necessarily demonstrate that "the principles underlying the procedures are valid." *People v. Collins*, 49 Misc.3d 595, 613, 15 N.Y.S.3d 564, 576 (Sup. Ct. 2015). Instead, the Court must independently determine that the internal validation is adequate to cover the procedure used in any particular case.

Here, the Army Lab had not internally validated STRmix to interpret mixtures containing DNA from more than four contributors. Exhibit F, DFSC DNA 114.1, Pages 6-10. Accordingly, the Army Lab would not have been able to analyze a mixture containing five or more contributors because its results would not have been deemed reliable.

The developers of STRmix have recently published multiple peer-reviewed studies concluding that in order to obtain reliable results with STRmix, laboratories must exercise great care when analyzing case scenarios, such as this one, in which first-order relatives of the person of interest likely contributed DNA to the samples in question. *See*, *e.g.*, Exhibit CC. These studies demonstrate that when multiple first-order relatives are likely present in a DNA mixture and when reference samples for all potential donors are unavailable, STRmix will produce unreliable results.

Here, no analyst could reliably conclude that the DNA samples to be analyzed contained fewer than five or six family members who lived in the household and used the parental bed. Importantly, the Army Lab's version of STRmix was not validated for analyzing DNA samples containing more than four contributors. Exhibit F, DFSC DNA 114.1, Page 9 ("6.6.3.2 Mixtures determined to be more than four individuals should not be interpreted further due to mixture complexity."). Moreover, STRmix has never been generally validated for interpreting mixtures containing more than five contributors. *Ortiz*, 2024 WL 2889873, at *22-30 (excluding expert DNA testimony where there were "numerous" indications that analyst had underestimated

number of contributors, giving rise to "a significant likelihood that [the number of contributors] was six or greater[,]" and holding, despite argument by government, that "there has not been general acceptance or peer review approval of STRmix for six-person mixtures"). The *Ortiz* Court distinguished several cases that have held that the number of contributors determination is a fact issue for the jury on the ground that none of those cases involved NOC determinations in six-person mixtures where potential relatives were involved. *Id.* at *22 (distinguishing *United States v. Barton*, No. 8:14-CR-496-T-17AEP, 2016 U.S. Dist. LEXIS 124485, 2016 WL 11469438, at *7 (M.D. Fla. Sept. 10, 2016) (three-person mixture); *People v. Burrus*, 81 Misc.3d 550, 200 N.Y.S.3d 655, 730 (Sup. Ct. 2023) (two or three); *State v. Warner*, No. A-15-858, 2016 Neb. App. LEXIS 161, 2016 WL 4443559, at *5 (Neb. Ct. App. Aug. 23, 2016) (two and three); *People v. Debraux*, 50 Misc.3d 247, 21 N.Y.S.3d 535, 542 (Sup. Ct. 2015) (same); *United States v. Morgan*, 53 F.Supp.3d 732, 746 (S.D.N.Y. 2014), aff'd, 675 F. App'x 53 (2d Cir. 2017) (two or three); *People v. Davis*, 75 Cal. App. 5th 694, 722, 290 Cal. Rptr. 3d 661 (2022) (three)). *Ortiz* further held that if the number of contributors was determined inaccurately, STRmix would produce unreliable results because STRmix had not been generally validated to interpret six-person mixtures. *Id.* at *23–30. The Court found that STRmix had not been validated to interpret six-person mixtures based on three peer-reviewed scientific articles: T. Moretti et al., *Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles*, 29 FORENSIC SCI. INT'L: GENETICS 126, 126–44 (2017), attached as Exhibit DD; J-A. Bright et al., *Internal validation of STRmix – A multi laboratory response to PCAST*, 34 FORENSIC SCI. INT'L: GENETICS 11 (2018), attached as Exhibit W; San Diego Police Department Forensic Biology Unit Validation of STRmix™ Software Addendum (2016), attached as Exhibit EE. *Id. Ortiz*'s reasoning applies persuasively on these facts.

Because there is no way to reliably determine that the stains tested in this case each contained DNA from fewer than five or six related contributors, the government cannot meet its burden under Rule 702 and *Daubert* of establishing that STRmix is reliable as applied to this case.

VIII. CONCLUSION

The Army Lab's internal procedures for DNA extraction and interpretation neither contemplated nor were validated for a situation that is as complicated and challenging as this case. As a result, Ms. Green did not exercise caution when analyzing samples that she knew potentially contained DNA from multiple first-order relatives. Instead, she made and stuck to several flawed assumptions, based on what she had been told the prosecution's case theory was, that ensured her analysis would be unreliable. And when it came to interpreting the data, her lab was not even equipped to test the most relevant hypotheses for this case, so she did not do so. For these and all the other reasons set forth above, Ms. Green's testimony should be excluded along with all other evidence relating to the DNA analysis performed in this case.

Dated: December 16, 2024
Brooklyn, New York

Respectfully submitted,
/s/
Samuel I. Jacobson
Marissa Sherman
Deirdre D. von Dornum
Austin Dean (Legal Intern)
Federal Defenders of New York
Counsel to Tyler Johnston

Cc: Counsel of Record (by ECF)
Tyler Johnston